数字人文平台个性化服务的功能设计研究
——以上海图书馆为例

■ 刘沛中 戴晴宜
上海图书馆上海科学技术情报研究所 上海 200031

摘 要：[目的/意义]基于用户个人的研究兴趣和需求方向，探索实现数字人文信息精准推送的实现路径，以优化我国数字人文平台的功能。[方法/过程]以上海图书馆的数字人文平台为例，探索知识图谱、用户画像以及热点推送三种服务方式在人文学科领域的应用，以提升现代图书馆服务质量，为用户提供更加高效优质的服务。[结果/结论]研究显示，将高效的人工智能推荐算法与数字人文服务相结合，能够为用户提供各种精准优质的推介服务，将高价值信息精准地推送到用户面前。

关键词：数字人文 知识图谱 图书馆服务
分类号：G250.7
DOI：10.13266/j.issn.0252-3116.2021.24.006

1 引言

随着数据科学、大数据、地理信息系统、文本挖掘、信息可视化等现代信息技术手段的飞速发展，数字人文（Digital Humanities，DH）以其跨学科、跨领域的特点，深刻地影响着历史地理、文学、计算机科学等学科，成为传统人文研究的有力补充和强劲动力。特别是在大数据环境下，世界各国的数字人文会议[1-2]、数字人文研究项目[3-4]、数字人文中心[5]均开展得如火如荼。

数字人文浪潮推动数字人文平台在国内外的广泛出现。数字人文平台是指为数字人文研究学者提供一系列数字人文学科领域相关的资源、数据、工具和服务的网络平台。这类平台通过收集和组织相关信息，实现对数字人文信息资源的统一访问，成为分析和传播数字人文信息的桥梁，满足人文学者研究需求的多样性，为使用技术推动跨学科的人文研究提供支持[6]。目前，国外数字人文平台有“美国历史协会（AHA）”数字人文平台、“ACO * HUM - 人文科学中的高级计算”数字人文平台等；国内的数字人文平台较少，如台湾大学“数位人文研究中心”和武汉大学“数字人文研究中心”已设立相应的数字人文平台。

当前数字人文平台大多只是使用关联数据技术对文献在元数据层面的知识进行组织和发布，对海量的数据缺乏足够的挖掘、提取和过滤，导致数字人文服务平台提供的服务价值密度低，用户无法及时高效的获取有价值的信息，造成资源浪费等问题。为了优化数字人文平台的服务，本文尝试以上海图书馆的数字人文平台实践为例，探讨综合运用知识图谱推荐[7]、用户画像推送、基于热度推荐等方式实现数字人文个性化服务的可能性与潜力。

2 数字人文平台的研究与发展现状

国内外图书馆界早已深入研究数字人文。南京大学信息管理学院的何盼盼根据国内相关文献的发表量进行统计后[8]指出国内研究可分为 3 个阶段：前期萌芽阶段（2011 - 2012 年），中文相关文献的出现比国外晚了近十年，第一篇为“第三届中美数字时代图书馆学情报学教育国际研讨会”上的一份发言稿[9]，该文献提出合格的“嵌入式馆员”应该具备数字人文、统计和计算方法、自然语言处理、语料库语言学等多方面的知识，并据此提供知识咨询服务；起步阶段（2013 - 2015 年），研究文献缓慢增长，年文献量保持在 5 - 15 篇；高速发展阶段（2016 年以后），2016 年北京大学图书馆启动数字人文系列活动，包括举办首届北京大学“数字人文”论坛、邀请国内外数字人文学者进行专场讲座等，故 2016 年学术论文数量呈“井喷式”增长。就近几年

作者简介：刘沛中，馆员，本科，E-mail：pzliu@libnet.sh.cn；戴晴宜，部门副主任，工程师，本科。
收稿日期：2021 - 05 - 06 修回日期：2021 - 10 - 18 本文起止页码：53 - 60 本文责任编辑：杜杏叶

的发展趋势而言,国内数字人文领域的相关研究已处于高速发展阶段。

数字人文平台是数字人文技术的应用,本质上是“基于数字人文技术构建的一个聚合数据资源、数据工具、数据服务的知识共享平台、内容开放平台”[10]。但在这方面的文献研究仍然较少。截至2021年10月17日,笔者以“数字人文平台”为关键词,对“中国知网”文献数据库中的全部中文文献进行篇名检索,检索结果仅有36条。这些文献谈到了数字人文平台可以实现的综合性服务功能,包括文本挖掘、可视化分析、场景模拟与复原、语料库利用等,但较少从用户视角出发,对个性化服务的实现方法和功能设计进行研究。

图书馆的视角下,个性化服务就是以用户为中心,在研究用户行为、兴趣、爱好、专业和习惯的基础上,根据用户的个性化需求而开展的信息服务[11]。对有具体研究需求的用户而言,较之全面而扁平化的信息服务,更渴求精准定位的个性化服务。具体来说,就是根据知识图谱共现、用户画像等技术,依据各种渠道对数字人文信息进行收集、整理和分类,向用户提供和推荐相关的数字人文信息,从而为用户提供精准的信息服务。从整体上说,数字人文个性化服务是将传统的被动服务模式转换为充分利用数字人文平台的各种资源,开展以满足用户个性化数字人文需求的全方位主动服务。虽然,这种平台构建思维已经在各种商用平台上被大规模使用,但在数字人文平台中很少被应用到。根据笔者调研的台湾大学“数位人文学术研究平台”和武汉大学“数字人文研究中心平台”来看,其平台主要都是通过用户的查询来进行数据的反馈,并没有采集和分析用户数据,并据此来进行个性化服务。因此将个性化服务功能加入到数字人文平台中,可以为使用者提供了更加富有针对性的服务。数字人文平台个性化服务的功能设计,是一个值得深入探索的方向。

3 数字人文平台个性化服务的功能设计

为了向读者提供优质的个性化服务,上海图书馆的数字人文平台在规划初期就从三个方面进行功能设计:分别是以知识图谱的方式来进行数据建构、采集用户行为数据进行用户画像,以及基于近期浏览数据进行热度推荐。此实践中,个性化服务具体体现在两个方面:①提高数字人文平台信息服务对用户需求的匹配度,尽可能过滤信息杂音,解决热度偏向问题;②提高数字人文平台信息服务对用户需求的预测精准度,从被动服务转向主动而定位精准的预测,及时推送迎合用户个性化兴趣的数字人文信息。

3.1 共现功能

“共现”是指文献的特征项描述的信息共同出现的现象,这里的特征项包括文献的外部特征如题名、作者、机构等,以及内部特征如人物关系、机构沿革、相关事件等。检索结果的共现需求是指上海图书馆数字人文平台根据用户的检索行为,除了以检索词匹配相关文献的外部特征外,同时根据文献的内部特征向用户推送与之有关联的文献知识[12],帮助平台揭示信息的内容关联和特征项所隐含的寓意。

3.2 推送功能

推送功能是指平台主动向用户进行信息、文献或是资源的推送。个性化的推送则是根据的用户的兴趣点,将不同的内容推送给不同的客户。个性化推送功能的后台机制是采集用户在平台使用过程中检索、浏览和留言的数据,进行分析处理形成用户画像数据,在将来用户使用平台的过程中主动向用户推送其兴趣相关的内容。

3.3 推荐功能

推荐功能是指向所有用户统一进行热点资源的推荐。上海图书馆数字人文平台需要收集用户群体近期检索和浏览的数据,通过分析用户与平台交互数据得到时事热点内容,在平台推荐栏展示给所有用户。推荐算法需要采用更加聚合的热度算法,解决热度偏向问题;平台需要提供热度排序表,把热点内容更全面具体的推荐给平台用户[13]。

4 数字人文平台个性化服务的功能实现

上海图书馆数字人文平台的个性化功能,可以运用不同的算法予以实现。本节主要对基于知识图谱嵌入与多神经网络的序列推荐算法[14-15]、用户画像提取算法和词汇热度算法进行介绍。

4.1 基于知识图谱实现共现功能

知识图谱中存放大量数字人文信息,结合知识图谱得到的共现信息,能向上海图书馆数字人文平台用户呈现更具体更全面的推荐。

4.1.1 知识图谱的表达形式

知识图谱(Knowledge Graph),是一种以图的形式存储人类知识的大规模语义网络。近年来,知识图谱在自然语言处理、问答系统、推荐系统等诸多领域取得了广泛且成功的应用。知识图谱中的节点表示实体,边表示关系。在知识图谱中,事实以三元组的形式进行表示。三元组一般记为<头实体,关系,尾实体>。图1为典籍知识图谱实例:

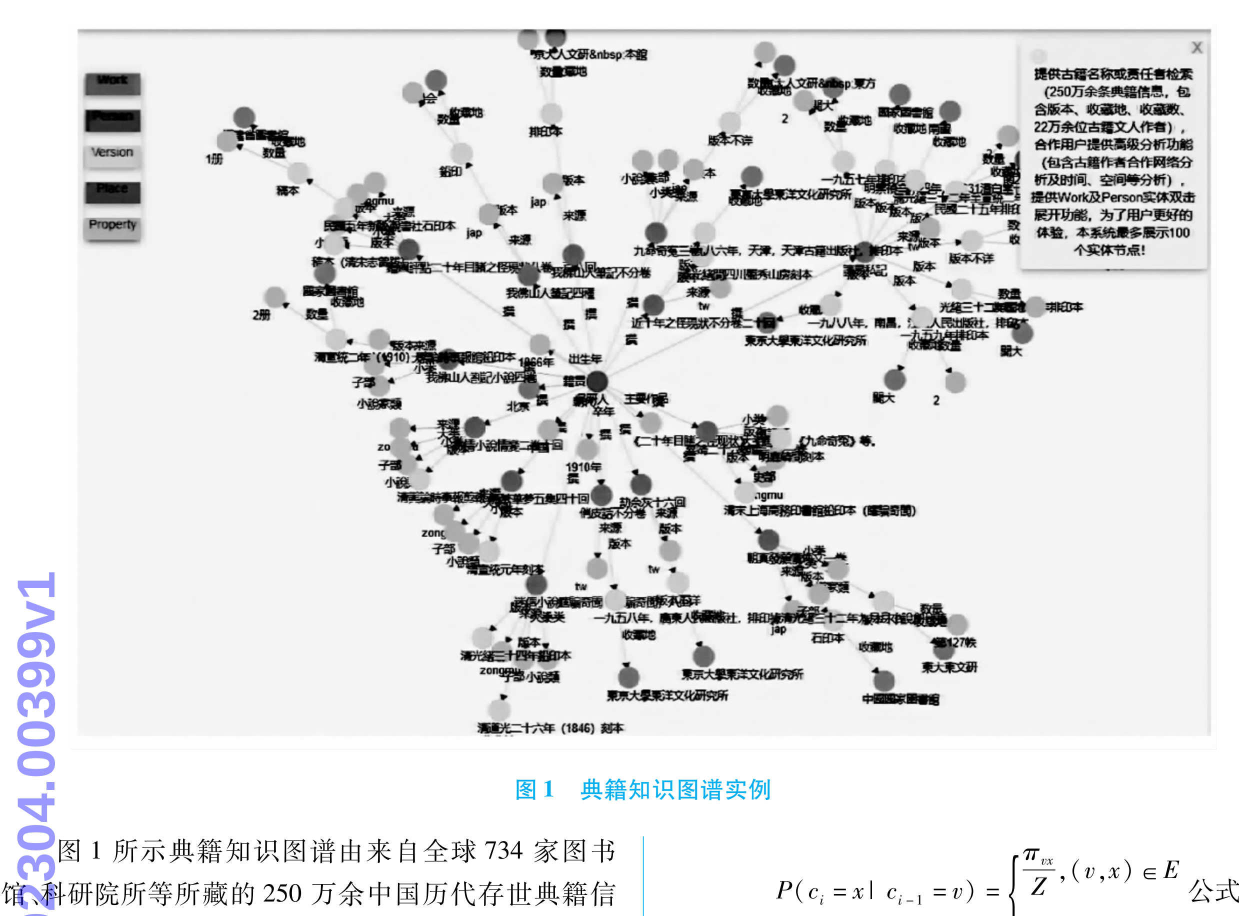


图1　典籍知识图谱实例

图1所示典籍知识图谱由来自全球734家图书馆、科研院所等所藏的250万余中国历代存世典籍信息组成，其中包含古籍实体649 549种(Work实例)、典籍责任者221 783位(Person实例)、古籍版本1 498 383个(Version实例)、地名节点13 960个(Place实例)，这四类节点及其之间的关系构成一个庞大的典籍知识图谱，节点、属性及边等形成了一个立体、多维、多用途的古籍知识关联网络，实现了对全球主要中国历代存世典籍书目信息较全面的描述，为研究者挖掘海量古籍书目数据背后隐藏的知识提供了一站式平台，大大增强了古籍知识服务功能[16]。

4.1.2　利用知识图谱抽取游走序列

node2vec[17]的思想是生成随机游走，对随机游走采样得到(节点，上下文)的组合，然后用处理词向量的方法对这样的组合建模得到网络节点的表示。

本文使用现有的知识图谱，然后使用node2vec算法思想构建随机游走序列。以上海图书馆数字人文平台为例，基于数字人文资源，结合事件、人、地名、机构，构建的知识图谱见图2。

使用node2vec中的随机游走方式采样随机游走路径，如(新文化运动，胡适，北京大学)等，然后从其中提取出数字人文实体序列。node2vec的广度遍历和深度遍历能很好地抽取实体间的同质性和同构性。随机游走的概率为：

$$P(c_i = x \mid c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z}, (v,x) \in E \\ 0, otherwise \end{cases} \quad \text{公式(1)}$$

其中，π_{vx}是未归一化概率，Z表示其中的归一化常数。c_i表示随机游走中的第i个节点，v和x表示知识图谱中的节点，E表示知识图谱。对于常见的随机游走，π_{vx}和实体边权重之间的关系为：$\pi_{vx} = \alpha_{pq}(t,x) \cdot \omega_{vx}$，$\omega_{vx}$为节点$v$和节点$x$之间的权重。系数$\alpha_{pq}(t,x)$计算方式如下所示：

$$\alpha_{pq}(t,x) = \begin{cases} \frac{1}{p}, d_{tx} = 0 \\ 1, d_{tx} = 1 \\ \frac{1}{q}, d_{tx} = 2 \end{cases} \quad \text{公式(2)}$$

其中，t表示上一个节点，x表示随机游走中下一个可能的节点，通过p和q的值来控制深度和广度游走，d_tx表示节点t和x之间的最短距离。本文使用*node2vec*的深度游走策略获取得项目序列作为下一步*item2vec*的输入，更好地捕获项目之间的相似性。通过知识图谱的序列抽取得到序列集合$H_k = \{I_1, I_2, \cdots, I_m\}$，其中$I_i = \{x_1, x_2, \cdots, x_n\}$表示生成的一条随机游走序列。

4.1.3　抽取用户行为的历史交互序列

假设令$U = \{u_1, u_2, \cdots, u_n\}$表示一组用户，而$I =$

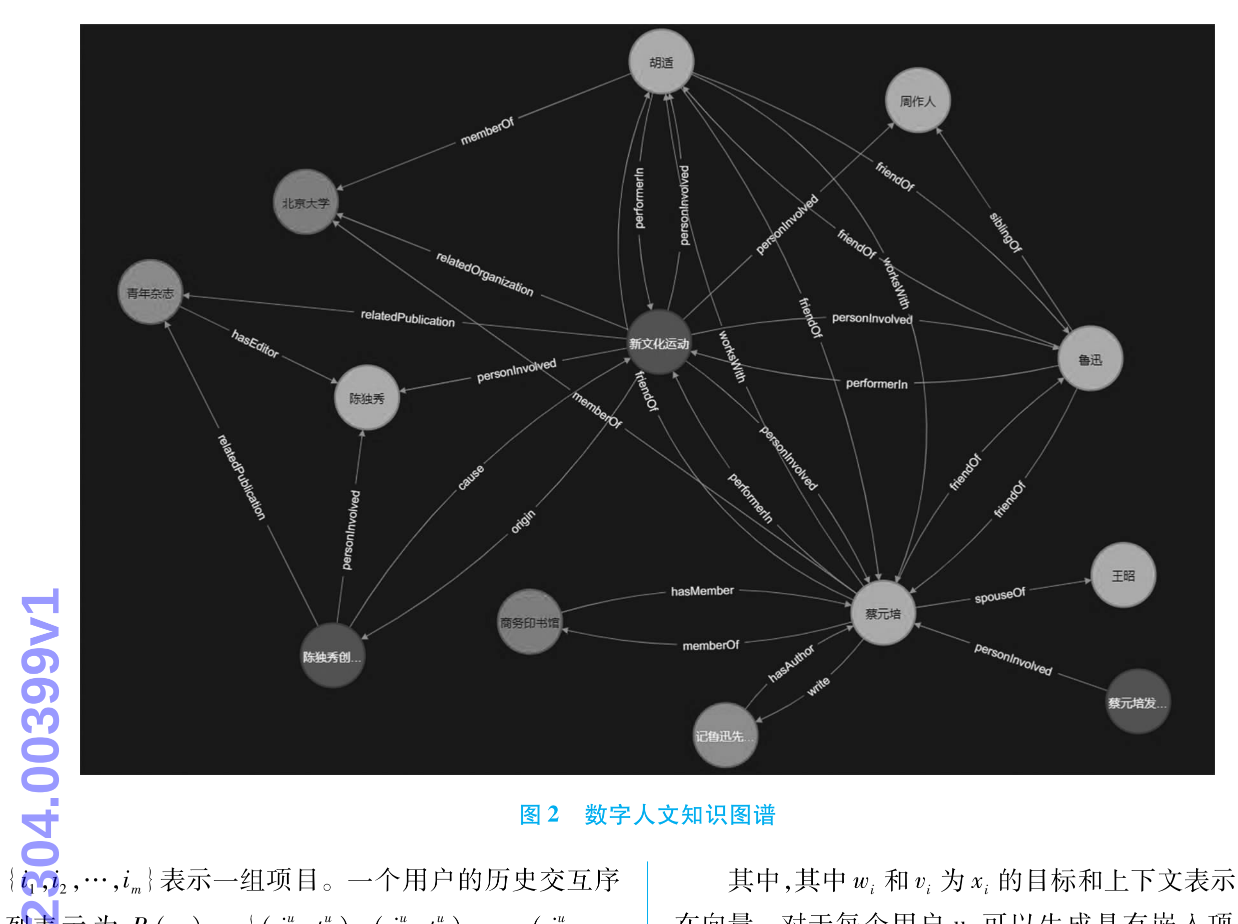


图 2 数字人文知识图谱

$\{i_1, i_2, \cdots, i_m\}$ 表示一组项目。一个用户的历史交互序列表示为 $B(u) = \{(i_1^u, t_1^u), (i_2^u, t_2^u), \cdots, (i_{|B(u)|}^u, t_{|B(u)|}^u)\}$，其中 (i_1^u, t_1^u) 表示用户 u 在 t_1 时间操作了项目 i_1。获取的 $B(u)$ 即为所需的历史交互序列。

4.1.4 将交互序列嵌入知识图谱向量

将从知识图谱抽取的序列与现有的用户交互序列相结合，作为 item2vec 的输入，最终得到项目的嵌入向量。知识图谱序列能够弥补采用项目序列嵌入较少考虑项目内容信息等缺点。神经项目嵌入模型类比于词向量模型，用户交互的项目随着时间自然地形成顺序序列，知识图谱得到项目序列，将两者结合类比于自然语句。同一作者、同一出版机构和拥有相同上下文信息的项目在嵌入空间上中离得很近。具体来说，给定用户交互序集合 $H = \{S_1, S_2, \cdots, S_N\}$ 以及知识图谱得到的序列结合 $H_k = \{I_1, I_2, \cdots, I_N\}$，item2vec 技术的 Skip-gram 模型旨在最大化以下目标：

$$\arg\max\ target = \frac{1}{K} \sum_{i=1}^{K} \sum_{j \neq i}^{K} logp(x_1 | x_2)$$

公式(3)

其中，K 是序列 S_i 和 I_i 的长度，x_i 表示序列中的项目，$p(x_1 | x_2)$ 定义为 softmax 函数：

$$p(x_1 \mid x_2) = \frac{\exp(w_i^T v_j)}{\sum_k \exp(w_i^T v_k)}$$

公式(4)

其中，其中 w_i 和 v_i 为 x_i 的目标和上下文表示的潜在向量。对于每个用户 u，可以生成具有嵌入项的交互序列，如公式(5)所示：

$$R_u = \{v_1, v_2, \cdots, v_n\}$$

公式(5)

其中，v_j 表示项目 x_1 的 d 维潜在向量。

上述算法过程中，item2vec 对项目进行聚类，捕获项目的相似性，并生成统一的项目表示空间，其中嵌入产生的向量可以解释项目的相似性和顺序关系，聚类在一起的项目具有强关联性。

4.1.5 node2vec 和 item2vec 算法的优势

node2vec 和 item2vec 算法根据用户的检索行为，通过知识图谱的随机游走和嵌入项的交互序列，可以有效地在杂乱无章的图数据库中进行节点筛选。在展开同样维度的节点信息时，能够获得最能贴近用户的节点内容。上海图书馆数字人文平台使用 node2vec 和 item2vec 算法后，不但实现了平台的共现功能，同时能够将共现信息更加精准化地推送给用户。

4.2 基于用户图像实现推送功能

用户画像是一种刻画用户信息模型的技术，在挖掘用户真实数据的基础上，提炼出用户兴趣点，呈现虚拟的用户信息全貌，以此为基础为用户提供精准个性化推送服务。本文通过所提出的序列建模框架建模用户的用户图像。

4.2.1 用户画像算法

本文将卷积神经网络(Convolutional Neural Network,CNN)作用于用户的兴趣点,充分考虑用户的历史交互序列上下文信息来学习用户的兴趣点,来形成用户画像。

兴趣的学习首先需要根据用户交互序列时间戳的信息划分用户的交互序列,然后使用 CNN 学习用户的兴趣点。本文将用户的序列分为长中短 3 个时期,动态地学习用户的偏好,卷积偏好建模的模型如图 3 所示:

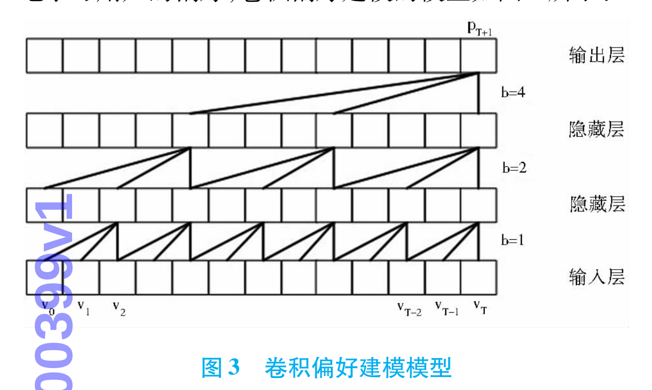


图 3　卷积偏好建模模型

本文使用一维卷积神经网络(CNN)学习用户的兴趣点和不同时间段的兴趣点。CNN 有良好的兼顾上下文信息的能力,能够很好地建模用户在一个时间段的兴趣点,挖掘用户的兴趣偏好。传统的 CNN 要获取充分的上下文信息,需要一个非常深的网络或者非常大的过滤器。本文将膨胀卷积应用到兴趣点偏好学习中,对于一维序列输入 $X \in R^N$ 和滤波器 $f:\{0,\cdots,k-1\} \in R$,其中,N 为输入的维度,对该序列的元素的卷积运算 F 定义为:

$$F(s)=(X*df)(s)=\sum_{i=0}^{k-1} f(i)\cdot X_{s-d\cdot i}$$

公式(6)

其中,d 表示项目向量维度,k 是卷积核大小,$s-d\cdot i$ 表示卷积操作过去的方向。f 表示卷积操作中的滤波器,s 表示序列中的元素。图 3 中 v_0 表示项目向量,P_{T+1} 表示第 $T+1$ 次输出结果,b 表示膨胀卷积的膨胀因子。因此,膨胀卷积在每 2 个相邻的滤波器之间引入一个固定的阶跃。当 $b=1$ 时,膨胀卷积变为规则卷积。使用大的膨胀因子可以使顶层的输出代表更大范围的输入,从而有效地扩展了 CNN 的接收范围。本文的残差块包含一个分支,该分支通过 F 的一系列变形,其输出被添加到块的输入中:

$$o=Activation(X+F(X))$$

公式(7)

通过 CNN 的学习,得到了用户 u 的兴趣点序列 $P_u=\{P_1,P_2,\cdots,P_j\}$,该序列就是用户 u 的个人画像数据,能够很好地刻画出此用户的喜好。

4.2.2 CNN 算法的优势

CNN 算法通过多次的卷积计算,从海量用户数据中提取出特征值。在上海图书馆数字人文平台中,先收集用户的检索行为和浏览内容,再通过 CNN 算法进行特征值提取,而提取出的特征值则代表了用户的兴趣点序列,也就是用户的个人画像数据。然后以这些个人画像数据为基础,在该用户再次进行检索的时候,将最为贴近此用户的检索结果优先呈现给用户,从而实现了用户画像的推送功能。在使用了 CNN 算法之后,上海图书馆数字人文平台可以更好地依据不同用户的兴趣点将对应的检索结果优先推送给用户,极大地改善了用户的使用体验。

4.3 基于热度算法实现推荐功能

将热点内容推荐给用户是时下的潮流所向,用户也希望开拓视野,获取热门、值得关心的数字人文内容。上海图书馆数字人文平台采用了热度算法,采集近期用户与平台的各项交互数据,实现了为用户提供热点资源推荐的功能。

4.3.1 基于时间线的人文词汇热度计算模型

定义:每个词汇 w_i 统计 4 个数据,在时间段 d_0 内,所有包含 w_i 的上海图书馆数字人文平台总搜索数 a,没有包含 w_i 的总搜索数 b。在 d_0 之外,所有包含 w_i 的搜索数 c,没有包含 w_i 的搜索数 d。如表 1 所示:

表 1　搜索统计

检索词 / 时间段	w_i	$\bar{w}_i$
$d_t \in d_0$	a	b
$d_t \notin d_0$	c	d

词汇热度的计算,如下述公式(8)所示:

$$h_{w_i}=\frac{(ad-bc)^2}{(a+b)(a+c)(b+c)(b+d)}$$

公式(8)

4.3.2 基于多用户集群的词汇热度算法

在上海图书馆数字人文平台中,使用层次聚类算法将用户按照兴趣点聚类为多个集群。层次聚类算法是根据每两个对象之间的距离,将距离最近的对象两两合并,合并后产生的新对象再进行两两合并,以此类推,直到所有对象合为一类。出于公平的原因,应对每个集群进行词汇热度的运算;否则,用户数多的集群在整个用户空间中占比大,对热门词汇的控制强。

首先并行在每个集群 p_i 中统计时间段 t_0 内的用户搜索各词汇的搜索次数,取搜索次数前十的词汇作为热门词汇,以第一个集群 p_1 为例,热门词汇集记为 T_1。然后,统计每个热门词 $w_i \in T_1$ 在集群 p_1 中时间段

t_0 内的所有包含 w_i 的总搜索数 a_1，没有包含 w_i 的总搜索数 b_1。在 t_0 之外，所有包含 w_i 的搜索数 c_1，没有包含 w_i 的搜索数 d_1，计算该热门词 w_i 的词汇热度。接着，将各集群的热门词汇的热度统计成一张热门词汇总表，相同的热门词汇热度值累加。最后，根据词汇热度进行排名，形成 Top-N 热门搜索列表。

4.3.3 词汇热度算法的优势

上海图书馆数字人文平台上使用的词汇热度算法综合使用了时间线和用户集群两个分支热度算法。时间线热度算法保证了词汇的热度必须随着时间流逝而衰减，从而防止某些历史热度值很高的词汇长期占据了推荐位置。而用户集群热度算法，则是在词汇热度计算之前先根据用户的兴趣点进行用户集群分类，然后在各个集群中分别计算词汇热度再相加形成最后的热度值。集群热度算法可以防止大集群对热门词汇控制过强，通过集群热度算法，得出的热度词汇会更加丰富，不会集中在某些特定领域。上海图书馆数字人文平台使用了上述两个热度算法后，呈现出丰富而多样热门推荐内容，达到了原始设计的预期要求。

5 数字人文平台个性化服务的实现效果

上海图书馆在 2020 年启动了“历史人文大数据平台项目”，整合了上海图书馆的各类数字资源，形成了目前的上海图书馆的数字人文平台。该平台将原有的数字资源进行知识重组，能够支持数据驱动的定量分析、可视化展示、文本分析、社会网络关系分析、地理空间分析等新的数字人文个性化服务。

个性化服务功能是上海图书馆数字人文平台的重要模块。使用知识图谱技术实现共现功能，使平台能够全面完整的呈现用户所查询信息的整体框架；使用用户画像技术帮助数字人文平台掌握用户的兴趣动态，给予用户最科学专业的个性化定制推送；使用热点算法技术使平台对整体用户数据的把控分析，向用户提供热点资源的推荐。本节将介绍上海图书馆数字人文平台在个性化服务功能在前台的页面呈现效果。

5.1 共现功能实现效果

在上海图书馆数字人文平台中，知识图谱共现的实现主要是用户检索关键字后的页面呈现，如图 4 所示。以“新文化运动”检索关键字为例，上海图书馆数字人文平台页面的左半边是基于外部特征进行文献检索的结果；依次呈现在右半边是根据检索关键字“新文化运动”和知识图谱游走序列生成的共现信息，其中包括人物、机构、相关事件和相关文献。

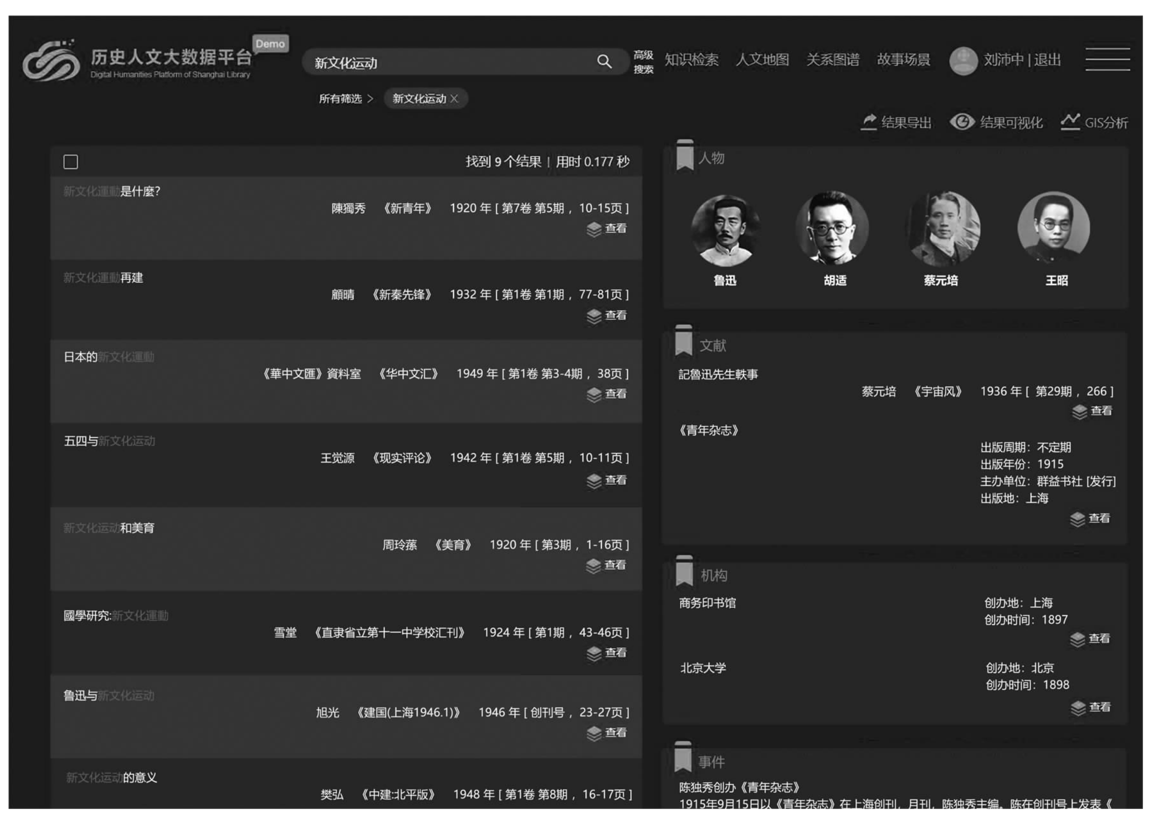


图 4　基于知识图谱实现检索共现的结果

刘沛中，戴晴宜. 数字人文平台个性化服务的功能设计研究——以上海图书馆为例[J]. 图书情报工作,2021,65(24):53－60.

5.2 推送功能实现效果

在上海图书馆数字人文平台中，结合用户画像，不同用户的关键字检索会呈现不同的检索结果，如图5所示，左右两侧是不同的用户采用同样的关键字进行检索，但是检索结果的呈现是不同的。左侧是新用户的检索结果，而右侧是老用户，平台根据老用户的检索历史、阅读内容和文献查阅方式形成的与平台的交互序列，让平台不断地更新该用户的用户画像，即更新其的兴趣点，最后导致检索结果的内容有所不同。

通过对图5右侧老用户的后台用户画像数据进行分析，发现右侧老用户先前经常检索的关键词为“五四”和“国民”，同时该用户经常阅读与“五四运动”相关的文献内容。因此该用户形成了特定的用户画像，在输入同样的检索词之后并产生大量检索结果的时候，系统将与该用户画像相匹配的数据优先推送给该用户。



图5 用户图像的实现结果图

5.3 推荐功能实现效果

在上海图书馆数字人文平台中，热门搜索在页面上的呈现为热度推荐榜单，如图6所示，根据热度算法计算得到的热度值将相关文献进行排序并展示。用户能够查看热度榜了解全体用户的兴趣所在，根据热度榜选择自己感兴趣的文献。



图6 热门搜索的实现

6 结语

图书馆与数字人文有着天然的联系，数字人文的发展给图书馆带来了机遇，更带来了挑战。本文针对数字人文平台的个性化服务进行分析，提出了将高效的人工智能算法与数字人文相结合，为上海图书馆的数字人文平台嵌入个性化功能，能够为读者提供各种精准优质的推介服务。通过对平台的测试与验证，结果基本满足预期要求。下一步，笔者将围绕如何构建更为丰富和智能的知识图谱，提高个性化服务的效率和精确度，优化读者阅读体验等方面继续展开研究。

参考文献：

[1] Centers. An international network of digital humanities centers[EB/OL]. [2021－04－26]. https://dhcenternet.org/centers.

[2] 朱本军，聂华. 跨界与融合：全球视野下的数字人文——首届北京大学“数字人文论坛”会议综述[J]. 大学图书馆学报，

2016, 34(5): 16 – 21.

[3] FITZELL S, LANGE L, LI-MADEO C. Linked Jazz: using linked open data to map community relationships inarchival documents [EB/OL]. [2021 – 04 – 26]. http://mysite. pratt. edu/ ~ infosils/SAAPosterAUG9. pdf.

[4] SULA C A. Digital humanities and libraries: a conceptual model [J]. Journal of library administration, 2013(53): 10 – 26.

[5] ACRL RESEARCH PLANNING AND REVIEW COMMITTEE. Top trends in academic libraries[EB/OL]. [2021 – 04 – 26]. https://crln. acrl. org/index. php/crlnews/article/ view/9137/10061.

[6] 程静. 国内外数字人文服务平台建设现状及思考[J]. 图书馆学研究,2020(2):37 – 42.

[7] 陈涛,刘炜,单蓉蓉,等. 知识图谱在数字人文中的应用研究[J]. 中国图书馆学报,2019,45(6):34 – 49.

[8] 何盼盼,陈雅. 我国图书馆数字人文研究述评[J]. 图书馆学刊,2018,40(4):132 – 137.

[9] SUSAN SCHREIBMAN, RAY SIEMENS, et al. A companion to digital humanities[M]. Malden: Blackwell,2004:16 – 27.

[10] 陈宇. 民国档案数字人文服务平台建设思路研究[J]. 北京档案,2021(8):28 – 31.

[11] 贾宏. 数字图书馆个性化服务技术述略[J]. 现代情报,2006(3):71 – 74.

[12] 董彦. 人工智能背景下图书馆知识服务创新研究[J]. 图书馆学刊,2019,41(7):78 – 81.

[13] 张磊,吾守尔·斯拉木,买买提依明·哈斯木,等. 大数据下微博推荐算法[J]. 激光杂志,2016,37(6):1 – 6.

[14] DONKERS T, LOEPP B, ZIEGLER J. Sequential user-based recurrent neural network recommendations[C]. Como: Proc of the 11th ACM conference on recommender systems. 2017:152 – 160.

[15] 沈冬东,汪海涛,姜瑛,等. 基于知识图谱嵌入与多神经网络的序列推荐算法[J]. 计算机工程与科学,2020,42(9):1661 – 1669.

[16] 欧阳剑,梁珠芳,任树怀. 大规模中国历代存世典籍知识图谱构建研究[J]. 图书情报工作,2021,65(5):126 – 135.

[17] GROVER A, LESKOVEC J. Node2vec: scalable feature learning for networks[C]//Proc of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. San Francisco: 2014:701 – 710.

作者贡献说明:

刘沛中:论文框架设计、数据调研和论文撰写;

戴晴宜:程序开发、测试。

Research on Personalized Services of Digital Humanities Platforms: Taking Shanghai Library as an Example

Liu Peizhong　Dai Qingyi
Shanghai Library, Shanghai Institute of Scientific & Technical Information, Shanghai 200031

Abstract: [**Purpose/significance**] In order to optimize the functions of Chinese digital humanities platforms, this article explores the path to realize the accurate push of digital humanities information based on users' personal research interests and need directions. [**Method/process**] Taking the digital humanities platform of Shanghai Library as an example, this article tried to use three service ways of knowledge graphs, user portraits and hotspot push in the field of humanities to enhance the quality of modern library services, and provide users with more efficient and high quality services. [**Result/conclusion**] The research shows that digital humanities services combined with efficient artificial intelligence recommendation algorithm can provide users with a variety of accurate and high-quality referral services, and accurately push high-value information to users.

Keywords: digital humanities　knowledge graph　library service